

AJ-A097 030

BOLT BERANEK AND NEWMAN INC CAMBRIDGE MA
PROCEDURAL SEMANTICS AS A THEORY OF MEANING.(U)
MAR 81 W A WOODS
BUN-4627

F/8 5/7

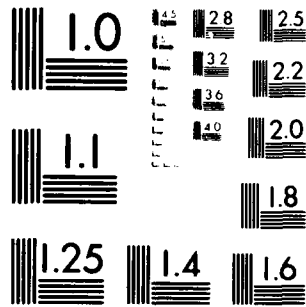
N00014-77-C-0371

NL

UNCLASSIFIED

1 of 1
25/07/80

END
DATE
FBI
5 81
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Bolt Beranek and Newman Inc.



LEVEL

12

Report No. 4627

AD A 097 630

Procedural Semantics as a Theory of Meaning

W.A. Woods

DTIC
APR 10 1981

March 1981

Prepared for:
Office of Naval Research

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution unlimited

81 4 10 036

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BBN Report No. 4627	2. GOVT ACCESSION NO. AD-4097630	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROCEDURAL SEMANTICS AS A THEORY OF MEANING.		5. TYPE OF REPORT & PERIOD COVERED Technical Report.
7. AUTHOR(s) W. A. Woods		6. PERFORMING ORG. REPORT NUMBER BBN Report No. 4627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Bolt Beranek and Newman Inc. 10 Moulton Street Cambridge, MA 02238		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0371 NSA 47-0
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBER
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE March 1981
		13. NUMBER OF PAGES 54
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Semantics, procedural semantics, theory of language, meaning, truth conditions, computational linguistics, computational semantics, semiotics, artificial intelligence, intelligent systems, knowledge representation, conceptual representation, meaning functions, recognition functions, abstract procedures, internal language.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report addresses fundamental issues of semantics for computational systems. The question at issue is "What is it that machines can have that would correspond to the knowledge of meanings that people have and that we seem to refer to by the ordinary language term 'meaning'?" The proposed answer is that the notion of truth-conditions can be explicated and made precise by identifying them with a particular kind of abstract procedure and that such procedures can serve as the meaning bearing cont'd.		

DD FORM 1 JAN 75 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract (cont'd.)

elements of a theory of semantics suitable for computer implementation. This theory, referred to as "procedural semantics", has been the basis of several successful computerized systems and is acquiring increasing interest among philosophers of language.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	
A	

Unclassified.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Report No. 4627

PROCEDURAL SEMANTICS AS A THEORY OF MEANING*

W. A. Woods

March 1981

Prepared by:

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

*Procedural Semantics as a Theory of Meaning, to appear in
A. Joshi, I. Sag, and B. Webber (eds.), *Computational Aspects
of Linguistic Structure and Discourse Setting*. Cambridge
University Press (in press).

TABLE OF CONTENTS

	Page
1 Introduction	1
2 Intelligent Systems	5
3 Internal and External Language	8
4 The Role of (External) Language	9
5 The Role of Sense Ambiguity	10
6 Acquisition of Meanings	12
7 Parametric Ambiguity	14
8 The Economic Necessity of Ambiguity	16
9 Semantic Interpretation	19
10 Semantics of the Internal Language	21
11 Motivations for a Procedural Account of Meanings	25
12 Some Inadequate Theories of Procedural Semantics	28
13 A Fallacious Criticism of the Procedural Approach	31
14 Partial Functions	34
15 Accessibility	37
16 Idealized Access	39
17 Knowing How versus Knowing Whether	40
18 Meaning Functions versus Recognition Functions	41
19 Abstract Procedures	44
20 Abstract Procedures as Intensional Objects	46
21 Towards a Theory of Meaning Functions	47
22 Conclusions	49

1 Introduction

For quite a few years now, a number of researchers (including myself) have been attempting to construct artificially intelligent machines that understand and use natural language. In doing so, a number of problems that have interested philosophers for centuries have suddenly taken on a new practical importance. Among these is the problem of meaning. If a machine is to understand questions and commands and take appropriate actions in response to them, then it needs some well-specified criteria for what those questions and commands mean.

In searching the philosophical literature for a suitable explication of meaning to use as a foundation for such a system, I found many useful insights, but no really adequate notion of what a "meaning" might be or how one might capture it in a computer. The closest notion that I could find was a concept of truth conditions that purported to characterize the circumstances under which a (declarative) sentence would be true and those in which it would be false [Carnap, 1964; Church, 1964]. However, these truth conditions are usually viewed as arbitrary functions from possible worlds or abstract models into truth values, following a tradition begun apparently by Frege [1892] and formalized and codified by Tarski [1944].

The Tarskian model theory provides a formal characterization of the circumstances in which a complex logical statement constructed from Boolean operators ("and", "or", "not") and universal and existential quantification would be true as a function of the truth values of their constituent elementary propositions (i.e., those not composed out of logical operators). The "circumstances" under which such expressions are either true

or false are defined by abstract "interpretations" (or "models") consisting of an assignment of a truth value to every possible application of a predicate to every individual in an assumed universe of individuals and predicates. No account is given of how such infinite assignments can be finitely represented or how they could be related to the actual state of the world.

In attempting to give a concrete and finitely representable explication to the notion of truth condition, I settled on the notion of a procedure as a familiar example of a finitely specifiable way to characterize an infinite set, and having the advantage of also permitting the definition of truth conditions for elementary propositions in terms of primitive operations of sensory perception. (The "sensory perception" of the machines that I first constructed were quite limited, but the theory was chosen with larger goals in mind.) This approach also had the advantage of providing a sensible characterization of the meanings of imperatives and questions as well as simple propositional assertions.

This approach, which I dubbed "procedural semantics" was developed in the context of a hypothetical airline information system [Woods, 1967, 1968], and received its first significant application in the LUNAR system [Woods et al., 1972], a system that answered natural English questions about the chemical analyses of the Apollo moon rocks. Approximately cotemporal with the development of LUNAR, Terry Winograd used very similar techniques to build his blocks world system, SHRDLU [Winograd, 1972], which simulates a robot moving blocks on a table in response to natural English commands.

Since then the concept of procedural semantics has generated

considerable interest and I fear some confusion. In this paper, I want to discuss some of the things that procedural semantics is not (at least to the extent that the term refers to an attempt to explicate the notion of "meaning"), to present a number of subtleties that must be dealt with if an adequate theory of meaning is to be obtained, and to outline what I feel are the beginnings of such a theory.

Studies of semantics have traditionally taken one of two forms - either the specification of semantic interpretations of ordinary language in some more or less well-defined notation, or the formal specification of the truth conditions or proof procedures for formal languages. However, the two activities are usually carried on independently, almost as if they were unrelated fields of study. When one attempts to design computer systems to understand natural language, one must not only attack both of these problems in a compatible way, but one must also address the basic problem of how the truth conditions of sentences relate to the sense experience of the computer. In addressing this problem, it is difficult not to be interpreted as espousing traditional theories of verificationism and reductionism. However, what I propose is significantly different from those traditions. Although verificationist theories are generally discredited, almost everyone concedes that truth conditions are somehow "related" to perceptions. However, almost no one seems willing to be specific about what this relationship is. The approach I will present here is the best one I have found to date that attempts to account for this relationship.

In reading what I am about to say about procedures as devices for specifying the relationship between truth conditions and perception, it is important to realize that I am not

espousing anything quite like any existing verificationist or reductionist theory. In particular, the abstract procedures that I advocate as devices for linking truth conditions to perceptions are not procedures that people can usually execute to determine the truth of a proposition, and they are certainly not intended to be substitutable for those propositions (so that the behavior of natural language semantics is simply reduced to the rules of computation of machines). Although I will take the computation of machines as a foundation on which to build up descriptions of the semantics of natural language, those descriptions in no way follow from the principles of computation, but are rather an independent theory embodied in a computational medium in much the same way that theories of many kinds can be expressed in the predicate calculus.

In general, the procedural semantics approach is a paradigm or a framework for developing and expressing theories of meaning, rather than being a theory of meaning itself. It is possible to formulate obviously false theories of meaning within this framework, as well as (I will argue) correct ones. I will demonstrate that some of the most direct applications of the idea of identifying the meanings of utterances with procedures are not in fact adequate as theories of meaning. In particular, it is not sufficient to identify the meaning of an utterance with whatever procedure a machine happens to execute in response to it, with the procedure used to make incremental changes to a machine's memory store, with the procedures used to carry out a command, or even necessarily with the procedures normally used to decide whether to believe an assertion to be true. All of these procedures are important for a language understanding system to have and for scientists to explicate, but they do not correspond to the thing to which we seem to refer when we use the ordinary language term "meaning".

In what follows, I shall be discussing problems that I believe hold equally for human beings and machines if they are to adequately use and understand natural language utterances (although I have encountered them in the context of designing artificially intelligent machines). That is, I will be discussing problems that must be faced by any system (human, machine, alien, or whatever) that is to deal with natural language utterances in something like the way that humans currently do, and which we ordinarily refer to as "understanding" and "appropriately using".

2 Intelligent Systems

I think it is useful to begin an investigation of the semantics of English with a fairly clear image of the role of language in relation to the overall mental activity of the system that uses it. I will begin with a brief outline of a theory of intelligence put forward by Daniel Dennett [1974], which I find highly satisfying and a useful precursor to a theory of the evolution of natural languages. Dennett presents a plausibly mechanistic account for an array of increasing levels of intelligence based on a view of hypothesis formation and testing as a kind of internal natural selection. Moreover, he presents an argument for this as a kind of mental organization that could itself plausibly result from natural selection. Specifically, he points out that, for sufficiently low order organisms, the behavioral characteristics of that organism in response to stimuli are essentially "wired in" by their genes, and their overall behavioral program is only altered by genetic selection over many generations of individuals. He points out

the evolutionary advantages for an organism to have an internal model of the world against which to test hypothesized courses of action prior to carrying them out. Specifically, such an internalized evaluation of hypothetical actions permits one's theories to die instead of oneself. The advantage of such intelligence to a species is the ability to adapt to a changing environment within the lifetime of a single individual.

An intelligent animal, by this account, would have a behavioral program that was not completely determined by its genes, but which was generated by some internalized behavioral program hypothesizer (or program modification hypothesizer), and evaluated by some internalized behavioral program evaluator. These program hypothesizers and evaluators could either be determined genetically, or could themselves be the result of internalized program hypothesizer hypothesizers, program hypothesizer evaluators, etc., giving rise to higher and higher levels of intelligence as the number of levels of internalized adaptation is increased. This account not only gives a plausible explanation of how complex intelligent behavior could evolve from understood mechanisms, but also nicely predicts such facts as the increasingly long period of maturation of children as a function of the level of intelligence of a species, the development of pathological patterns of behavior as a result of exposure to certain kinds of environments during maturation, and a limitation on the degree of intelligence that can be obtained in this way (i.e., by repeated iterations of selecting the selectors, etc.) by virtue of the limited lifetimes of the individual.

Interestingly, from my point of view, Dennett's account also motivates the evolutionary potential for culture and natural language. That is, given this perspective, it is only a short

step beyond Dennett's account to see the development of a language for communication as a means for some of this rapidly acquired adaptation to transcend the lifetime of a single individual, thus saving each succeeding generation from having to learn everything from scratch. A culture then becomes a kind of higher level organism with a collective memory and can be viewed as a means of escaping the limited lifetime of a single individual. (It is interesting to speculate from this point of view whether cultures themselves have limited lifetimes and whether ours is anywhere near maturity.)

From this point of view, we see that a particularly useful component of the behavioral program evaluator is a model of the external world that can be used to predict the outcome of actions without actually having to carry them out. We will assume that at least in humans, almost certainly in many monkeys and apes, and probably to a lesser extent in most mammals, there is a genetically determined program either for acquiring such a model or for producing a program to acquire such a model. (The manifestations of this program in our behavior are generally referred to as "curiosity".)

In order to internally store such a model, an intelligent system will need to have some internalized "notation" or "representation" for expressing believed and hypothesized facts about the world. (I hesitate to use the term "language" at this point because of connotations of similarity to a surface language that I deliberately want to avoid. Specifically, I do not want to invite attributions to this internal notation of properties such as temporal word order, structural or semantic ambiguities, situationally dependent meaning, composition out of English words, discrete rather than analog representation, or anything

leading to assumptions of homunculi inside the head that are interpreting the internal notations in ways analogous to the way we understand English. However, subject to these qualifications, I will subsequently refer to this as the "internal language").

3 Internal and External Language

The thrust of the preceding discussion of intelligence, which may at first seem somewhat of a digression, is to motivate a notion of an internal language that both logically and temporally precedes the development of external language, and in terms of which the meanings of external language expressions are defined. This internal language itself requires a semantics, and I would maintain that without some understanding of its semantics, one cannot have a complete semantic account of an external language. I will argue that this internal language is capable of vastly more discriminative subtlety than one's external language, which has evolved to a suitable compromise between the need for discrimination, and the economic costs of inventing and learning a large vocabulary and inventory of syntactic conventions, and of composing, articulating, and understanding long external sentences. As evidence for this, consider the ability of this internal language to store the criteria for discriminating a particular familiar face versus the apparent inability to convey such criteria by means of the external language.

In general, I will argue that human communication relies in a critical way on an ability of the receiver to deduce a much more precise understanding of the intended meaning of an

utterance than is conveyed by the words alone and the syntactic structure in which they are incorporated. Moreover, I will argue that this is not just an unfortunate characteristic of natural language, but is in fact an economic solution to the problem of communication in a situation in which it is not economic to develop external terms and conventions for all of the discriminations that one can make internally.

4 The Role of (External) Language

If intelligence was evolved by natural selection and Dennett's account is correct, then one should find the development of the capability for external language evolved in a similar way. Moreover, one should find that the capacity for external language is selected to support the kinds of transferral of world models and behavioral strategies from individual to individual in the way that I have outlined. One might then expect that other "higher level" uses of language, such as poetry, song, etc. would derive in some natural way from this basic need (possibly as a result of a higher level evaluator which values skill in the use of language in general, as an indirect way of achieving the more specific goal of world model transference.)

Although there are many different kinds of use of language, I will assume in the rest of this paper that its use for factual communication, explanation of principles, and complex instructions for behavior are the primary uses which motivate its evolution. Although I will be concerned specifically with the "literal meanings" of utterances, as opposed to the various

things that a speaker might intend to convey by an utterance (threats, irony, etc.), nothing I say will be inconsistent with an account of the pragmatic interpretation of speech acts in contexts other than literal factual communication. I will maintain, in fact, that an understanding of the literal semantics of an utterance is often essential to determining its intended pragmatic meaning. For example, the correct understanding of a statement "It's a lovely day" uttered when it is in fact pouring rain requires the knowledge that the sentence is not literally true.

With these preliminary caveats in hand, let me now devote my attention to the interpretation of literal meanings, and primarily to the interpretation of factual assertions.

5 The Role of Sense Ambiguity

In my account of the semantics of external language, I will make heavy use of the notion of ambiguity, by which I mean the presence of alternative distinct interpretations in the internal language that the speaker may have intended to convey, and among which the hearer must in some sense try to choose the meaning that the speaker intended. I will thus demand of my internal language that it have the capability to represent explicitly and unambiguously any and all of the distinctions that one is capable of making in resolving an ambiguity in a surface utterance, as well as being able to represent any of the more abstract (or vague) interpretations that one can make. I will assume that what is understood in response to most (but not necessarily all) utterances is something that is much more precise and less

ambiguous than that utterance in isolation would permit. The additional information used to determine this more precise understanding comes from the current values of indexical expressions such as "here", "now", and "current speaker", from knowledge of the world, and from knowledge of various speech act conventions, rules of the language game, emotional state of the speaker, etc. Whereas the determination of the current interpretation of the speaker's intended meaning will be considered fallible and in some sense "fuzzy", the meaning of each possible sense of the utterance will be assumed to be relatively precise.

I claim that from this point of view most words in English are highly ambiguous and that the intended sense is selected by context. As an illustration, consider the classical example "bachelor", which by the conventional wisdom has a number of senses including certain kinds of fur seals and baccalaureate degrees [Katz & Fodor, 1964], but is usually considered to have only one sense as an "unmarried adult male human". However, as Winograd points out [Winograd, 1976], even this sense has many different subtle shadings, including whether the person referred to "lives the life-style of a bachelor", is eligible for marriage, comes from a culture that permits several wives but has not yet filled his quota, etc. Thus, what we take someone to mean by a sentence (in Winograd's example, "Do you know any nice bachelors I could invite [to my party]") is characterized by uncertainty. I will claim that this is due to a fundamental parsimony of the external language vis-a-vis the internal language (to be discussed further shortly), and that each of the distinguishable senses must be explicitly expressible in the internal language. Notice that this account does not necessitate any fundamental assumptions of fuzziness of meaning of the internal language.

By contrast, consider Winograd's account of this example, which rests on the notion of a prototype or "exemplar" of the general concept of bachelor to which various individuals can be matched with varying degrees of closeness of match. By this account, the concept of bachelor itself is a fuzzy concept in the sense of Zadeh [1974], in which different individual instances satisfy different numbers of the features associated with the exemplar. This account does not explain the fact that when a person is asked the above question, something fairly precise is understood without necessarily attempting to match any particular individuals to the concept. Understanding this sentence correctly, it seems to me, requires a particular sense of the word (defined perhaps by a particular allocation of importance to the features from Winograd's list) to be selected as the intended sense. One does not just understand by this sentence some vague notion of an approximate match to the exemplar that is fundamentally the same regardless of context. Rather one must characterize in some way what match criteria must be met, or at least which criteria are more important. Some residual ambiguity may remain as to whether the host(ess) specifically wanted "eligible" males that are likely prospects for marriage, or merely men who are living the lifestyle of a bachelor, but such ambiguity can be kept distinct from a fuzziness in knowing exactly what the different senses mean.

6 Acquisition of Meanings

Just as what we take someone to mean by a sentence is characterized by uncertainty, so is the process of learning what people mean by a term from the occasions of its use and/or its

"definitions" in dictionaries. It may be possible to formulate quite precise and well-defined concepts in the internal language with which to try to model the use of a term as we have been exposed to it; the problem in language acquisition (with respect to the meanings of terms) is to discover which such concept or concepts to take as the meaning(s) of a given term. Once again, our uncertainty in knowing the meaning of a term as other people use it does not necessitate any fundamental fuzziness in the meaning of the concepts of the internal language.

The general state of someone in the process of learning the meaning of an external language term (and I believe this to be the normal state for many if not most of the terms even in the vocabulary of an adult) must consist in holding some hypotheses about the meaning of that term as others use it. Such a hypothesis could be expressed by an explicit set of (fully specified) internal concepts, one or more of which is hypothesized to be what other people mean by the term. Alternatively, one could express such a hypothesis by a collection of ("meta") statements I.e., metastatements in the sense of statements in a metalanguage about the language in question. These statements about the hypothesized meaning. Such a characterization would implicitly determine a set of possible internal concepts, but would not require their explicit formulation. In this mode it would be possible to express hypotheses that characterize an infinite class of possible meanings, and to express beliefs about the distributions of likelihoods of those hypothesized possible meanings being correct. (It would also be possible to express inconsistent hypotheses that could not be satisfied by any well-defined concept - and it is almost certainly the case that people can do this and do so without realizing it. I think this is especially

likely to happen when one coins a term for a concept in scientific theory formation and in philosophy.)

The formulation of metahypotheses about the meanings of a term also permits the expression of relative likelihoods or likelihood distributions for the different hypotheses, reflecting their degrees of confirmation or disconfirmation. Again, this does not entail any fundamental assumption of fuzziness in the semantics of the internal language expressions, although it does imply a rather complex process for determining the meaning to take for a sentence involving the use of a term for which one only has hypothetical metastatements to constrain its meaning. In the worst case, one may only be able to deduce hypothetical metastatements about the meaning of such a sentence.

7 Parametric Ambiguity

If one is to use the notion of ambiguity to account for the uncertainty in the meanings of terms as people use them, then in addition to the usual notion of (discrete) ambiguity, it seems necessary to introduce a notion of ambiguity that is continuously variable over a potentially infinite range of possibilities. This seems necessary, for example, for an account of the ambiguity of various "measure" predicates, such as "tall" and "contain" (i.e., predicates that can take qualifiers of amount and admit questions such as "How tall is John?" or "How much does it contain?"). When such terms are used in simple declarative assertions, they seem to assert that the appropriate measure (e.g., height) exceeds some threshold. I will treat such sentences as ambiguous with respect to this threshold, and will refer to such cases as "parametrically ambiguous".

The traditional account of such sentences as "John is tall" is to consider them discretely ambiguous with respect to various classes to which John might belong, taking the necessary threshold from the normative height for each class (e.g., "John is tall for a fifteen year old."). However, this account fails to explain a residual ambiguity in what counts as a significant deviation from the norm. That is, whatever threshold one picks for the normative height, there remains some intended scale of what counts as a significant difference that must also be inferred in order to know what the speaker intended to say. One assumes that one millimeter over the normative height would not count as "tall" nor would an additional meter of height be required (for humans - but not for buildings). Somewhere in between is a level of discrimination that the speaker was intending his hearer to assume, and we need to infer some value of this parameter in order to fully determine what assertion the speaker was making (e.g., what fact about the world he wanted the hearer to believe).

It is usually not critical in practice to get this value precisely right, but some value needs to be assumed. In many cases, the utterance of such a sentence about a person whose height is known to the hearer is in fact used to "calibrate" the speaker with respect to this ambiguity - i.e., to specify what the speaker considers tall. In other cases, what the hearer presumably encodes in response to such an assertion is a specification of a range of parametric ambiguity that (s)he is unable to resolve, pending some future situation in which it may be further resolved or may make a difference. In general, it seems to be psychologically difficult for a hearer to maintain more than a few discrete ambiguous interpretations of an input utterance, and one can easily imagine how this is a costly

operation for a mechanical system. For a parametric ambiguity, however, the encoding of the ambiguity requires only the lack of choice of a value for the parameter or some description of a range of possible values (e.g., the end points of an interval), perhaps accompanied by some description of a likelihood distribution for the possible values. One would therefore not expect the same level of difficulty for holding open this kind of ambiguity.

8 The Economic Necessity of Ambiguity

When one thinks of it from the perceptual point of view, a linguistic speech act is merely another kind of perceived entity. Moreover, the structural decomposition of a sentence into its individual constituents and their relationships to each other is an intimate part of its perception as a sentence. (Notice that ambiguity enters the picture here as an aspect of perception in exactly the same way that certain diagrams and visual scenes are ambiguous and can be perceived in more than one way.) When viewed in this way, it becomes clear that the perceptual complexity of words and constructions (as spoken waveforms or patterns of ink on paper) is comparable in some sense to the complexity of other perceived entities such as boxes, people, automobiles, etc.

From the preceding discussion, it follows that characterizing the perceptual conditions for some concept, plus also characterizing the perceptual conditions for some word that is uniquely associated with that concept, is approximately twice as much effort as representing the perceptual conditions for the

concept alone. Specifically, the process of forming a concept by the modification or combination of other concepts is comparable in complexity to the process of forming a lexical concept by the modification of some existing word or the composition of phonetic elements, syllables, or some such constituents. Moreover, the process of forming a named concept requires not only this doubling of complexity in what is stored, but also the effort involved in coining and remembering a suitable external name - a process that involves appropriate problem solving tasks such as generating candidate names and testing them for memorability, potential misleading associations, previous use for other concepts, etc.

Given the situation just presented, it should not be surprising that in general one has many more concepts than one has external names, or that one habitually uses the same external name for many different but related concepts (especially where confusion is not likely to result, but also in places where it is). As an example of the former, consider the concept of the small cylindrical projection on the end of a shoe lace that keeps it from unraveling and makes it easier to thread through holes. I expect that you have little difficulty in identifying the concept that I have in mind, although the above description falls far short of completely specifying the concept to someone who has never seen a shoelace. (In a more specific context, such as while displaying a shoelace without one, I would be able to get the point across with as little as "the end is missing", which given nature of strings it would be absurd to take literally.)

The multiple use of the same word for different concepts is so frequently done and the nuances of meaning shift are sometimes so subtle that we are largely unaware that it is happening. Only

the more glaring examples that are listed in dictionaries as multiple senses seem to make their presence known. However, there are a number of very subtle, systematic uses that are not normally listed as separate word senses in dictionaries. One of these is the use of a word both to refer to an object and to refer to a representation of that object in a picture or diagram (highlighted in such ambiguous phrases as "painting nudes").

This principle of economy also predicts the use of regular systematic devices for coining names for concepts without having to coin completely new words. Examples are the frequent use of a noun to name an action that is related to it in some way (e.g., "to land a plane", "to dock a ship", "to chair a meeting", etc.) and the highly productive use of so-called noun-noun modifiers to form multi-word names for concepts (e.g., "dog house", "pot cover", "ambiguity tester", "idle-speed adjusting screw", "eight-track stereo cartridge tape player", etc.).

In a similar way, the same economy principle motivates the use of devices such as anaphoric reference, ellipsis, and other locutions involving a degree of lack of specificity. That is, one must assume that the internal language has little difficulty in making repeated references to an internal concept by some unique "handle" or pointer, in much the same way as if it had a unique proper name for that concept. The external language, however, has no such simple facility, and uses instead such devices as pronouns and other anaphoric expressions. The use of deixis, ellipsis, and various other such techniques are all motivated by the differential economics of having something in mind, versus expressing it explicitly in the external language.

In summary, the thrust of previous arguments is that, unlike

the claim of Montague [1973, 1970] that one could in principle give an adequate semantic theory of a natural language directly, as if it were a formal language, I maintain that natural languages have an essential difference from formal languages that necessitates a two-stage account of their semantics. The first stage must cope with the ambiguity inherent in natural languages, while the second stage consists of giving a formal semantics for the underlying meaning representation in which the alternative senses of ambiguous sentences are expressed. In fact, even Montague's accounts involve an intermediate entity (an analysis tree) that serves this disambiguation function.

9 Semantic Interpretation

Given the hypothesis of an internal language in which the interpretation of English sentences are expressed, the problem of semantic characterization of English becomes a two-stage process. The first stage consists of assigning to an input sentence one or more possible interpretations in the internal language, while the second consists of characterizing the semantics of the internal language. Historically, the first stage, which we will call semantic interpretation, has been the domain of most concern to the linguistic semanticist, while the second is traditionally of more concern to the philosopher. I will argue that both rules for assigning interpretations to sentences and the characterization of the meanings of the internal language expressions can be modeled by means of procedures. However, the nature of the two kinds of procedures and the way that they are used are quite different.

Semantic interpretation procedures as defined above are essentially translation procedures that are executed as part of the understanding of a sentence to produce a set of possible interpretations of the sentence. This process can be modeled by a variety of automata such as ATN grammars [Woods, 1970] or systems of rewrite rules as in a Transformational Grammar [Chomsky, 1965]. Woods [1978a] gives an exposition of how such translation is accomplished by means of formal semantic interpretation rules in the LUNAR system.

As I have pointed out, the sentences in the external language are in general highly ambiguous, although many different factors can be invoked by the understanding system to attempt to determine which meaning was intended. It seems that, at least conceptually, the performance of semantic interpretation involves the use of a fairly regular system of conventionalized literal semantic interpretation rules which determine a range of possible literal meanings, from which a variety of pragmatic considerations determine the intended meaning of the utterance in context. Sometimes this latter is done by simply selecting from among the possible literal meanings, but often it is done by the further application of various more or less conventionalized pragmatic or "speech act" rules.

I say the above organization is the conceptual organization in order to avoid the implications that there is necessarily an actual serial process in which the alternative semantic interpretations are explicitly enumerated before the pragmatic rules operate. In particular, one may want pragmatic rules to operate in the selection of which semantic interpretation rules to apply, or to interact with the semantic interpretation process in some other way. In most cases, however, a literal

interpretation (or parts of one) appears to play an important role in the application of speech act rules. For example, in interpreting certain kinds of ironic statements, it is necessary to conclude that the speaker is not intending to express the literal meaning, but rather its opposite.

Although there is a great deal of work required to characterize both the conventionalized semantic rules and the conventionalized pragmatic rules (as well as a range of non-conventionalized problem-solving activities that are often used in determining intended meaning), there is no difficulty in principle in knowing what kinds of things such rules are. That is, they can be characterized as formal translation rules of some sort that can be expressed as computer programs or any of a variety of abstract automata. In particular, they can be embodied in an Augmented Transition Network grammar. When one turns to the semantics of the internal language, however, things are not quite so clear cut.

10 Semantics of the Internal Language

In the remainder of this paper, I will be concerned with the specification of the semantics of the internal language (in terms of which the semantics of external sentences are to be defined). Our previous discussion of intelligence suggests that a principal use of this internal language is to develop a taxonomy of the kinds of objects, events, situations, etc. that can occur in the external world and to formulate hypotheses about the properties of such entities and the cause and effect relationships among them. It is almost certainly necessary that this internal

language be extensible in the sense that new concepts can be created out of old ones, and then used as components of the specification of still newer concepts. That is, one should not visualize the internal language as having a closed vocabulary.

What must be primitively present in the internal language are some basic concepts and an inventory of fundamental operators that can be used to construct new concepts from old ones. Moreover, principles of economy (and perhaps also logical necessity) dictate that such new concepts must be usable as elements in the internal language in much the same way that the basic concepts are used. They should not be visualized simply as abbreviations for more complex expressions that are to be substituted for them. The utility of creating new concepts in a taxonomy comes from thereafter being able to use the new concepts as a single entity rather than copying out its definition on every use. (The distinction I am trying to make here is similar to the way that "closed" subroutines in a computer program may be "used" in several places although only one copy of the subroutine is explicitly stored. "Open" subroutines or "macros" on the other hand are "expanded in place", resulting in a separate copy of the macro definition for each use.)

It is important for the internal language to be able to develop concepts that express very precise and subtle distinctions between classes of objects and situations. A man's life in the jungle (or, for that matter, an ape's) can depend on the ability to distinguish subtle differences between otherwise similar plants and animals, some of which are edible and others dangerous. Moreover, by the above account, what makes this discrimination ability advantageous to an intelligent animal is the ability to associate cause-effect (or "if-then") predictions

with the various concepts in the taxonomy (e.g., "if you eat this, you will get sick" or "if you eat this it will satisfy your hunger"). The two most important aspects of this internal representational system would thus seem to be its ability to specify precise perceptual conditions for a concept and to characterize the "if-then" associations among such concepts (including concepts of various internal states such as hunger and sickness).

Although in what follows I will be primarily concerned with the semantics of this internal language, rather than its representational conventions and capabilities, the kinds of representations that I will be assuming are something like those of Ron Brachman's "structured inheritance networks" [Brachman, 1978, 1979; Woods & Brachman, 1978], within which is represented what I have been calling a "taxonomic lattice" [Woods, 1978b], an organization of concepts of different levels of generality in which the subsumption relationships and correspondences of conceptual subparts between concepts are explicitly represented. This lattice of concepts is used as a "conceptual coat rack" on which to hang various hypotheses and conclusions about different classes of entity.

Brachman's structured inheritance networks are a generalization of Quillian's notions of semantic networks [Quillian, 1966, 1968, 1969; Bell & Quillian, 1971] in which, among other things, generic information about concepts can be stored at its most general level of applicability and "inherited" by more specific concepts below (e.g., information about physical objects can be stored at a very high level, while specific information about rocks, animals, birds, etc. is stored at more specific levels). Economy principles of memory organization

suggest some such hierarchical or network organization in which information common to many concepts is shared.

The most salient feature of Brachman's networks is their treatment of inheritance for complex structured entities. These networks explicitly represent not only the subsumption relationship between two concepts, but also the correspondence of parts, attributes, and properties between them. Space does not permit a full exposition of these structures here, but the details of the representations are not critical to anything that follows. The use of such representations, in which given pieces of information are shared or inherited by many different concepts, contributes to the economic efficiency of an intelligent system, and answers a possible objection to my account due to apparent combinatorial problems in storing the necessary information, but otherwise is not essential to a discussion of what the concepts in the internal world model mean.

In the arguments that follow, I will not present any details of representation suitable for the internal language, but rather use a mixture of English, traditional predicate calculus notations, and some programming concepts from the language LISP [Berkeley & Bobrow, 1964; Weissman, 1967], which I hope will be relatively clear. For this purpose, it will be useful to know that the syntax of the LISP language consists of expressions in so-called "Cambridge Polish" notation, in which a functional operator appears to the left of its operands and the whole is enclosed in parentheses to indicate the grouping of function with arguments. (E.g., "(PLUS 2 3)" is the LISP notation equivalent to the "infix" notation $2+3$, and the conventional notation $f(x)$ would be represented as "(f x)" in LISP.)

11 Motivations for a Procedural Account of Meanings

An interpretation procedure that translates surface sentences into representations in an internal language is clearly procedural. However, it is not necessarily "semantic". In order for such a translation process to have anything to do with semantics, it must be the case that the semantics of the resulting representation are understood. This latter must eventually be specified by something other than just another such translation process. Procedural Semantics, at least as I use the term, refers not to the translation process of semantic interpretation, but to the use of procedures to characterize such things as truth conditions (for propositional assertions), conditions of satisfactory response (for imperatives and questions), and conditions of appropriate use (for various social speech acts, etc.).

While it seems clear that truth conditions for a term can be said to define its semantics, it is not necessary (at least it doesn't seem to be necessary) to interpret them as procedures. However, there are a number of motivations for using procedures as an explication of the notion of truth conditions, some of which were mentioned in the introduction. They are discussed more fully in Woods [1967]. While most of the arguments have to do with issues of efficiency or methodological advantages such as concreteness and clarity of understanding, the following argument seems to justify the approach for an account of how people work:

If a person (fully) understands a term characterizing a class of entity and is presented with a clear instance of a member of that class, he can recognize it and say "yes" and likewise he can reject a clear non-member. If we believe that in

doing this the brain is functioning as some kind of physical/electrical machine, then clearly there is some procedure that is being executed that recognizes members of the class and rejects non-members. If we do not recognize this process as a physical/electrical one, but rather the product of some uniquely mental "stuff", we may not necessarily be forced to admit the existence of such a procedure, but neither are we blocked from it. However, to avoid it would seem to require some direct and unexplained form of "knowing" independent of the functioning of the brain, the eyes, the visual cortex, and other sense organs. While I can't directly rule out such an account, it seems implausible and somewhat sterile as an account of ordinary human behavior. Therefore, a procedural account would appear to have significant face validity.

One of the major methodological advantages of the procedural semantics approach to meaning is that one does not encounter as many dilemmas when confronting things whose meanings do not seem to characterize truth values (things such as commands, promises, measure predicates, etc.). That is, the kinds of things that can be constructed out of the basic procedural elements of a universal machine constitute a richer inventory of conceptual entities than those that can be constructed out of the primitives AND, OR, NOT, and universal and existential quantification (at least without embedding a different conceptual system within a first order predicate calculus and specifying a set of axioms to characterize that system.)

Moreover, the procedural primitives of a higher level programming language such as LISP provide what seems to be a more useful basic set of operations out of which to construct potential "meaning functions" (by which I mean the functions that

define the truth conditions of propositions, satisfaction conditions for imperatives, etc.) For example, the procedural paradigm permits one to construct "primitives" for measuring such things as strength of pattern matching, statistical correlations, and weighted sums out of the same basic procedural primitives that one uses to characterize the meanings of the logical terms AND, OR, and EVERY.

Of course, one still has to determine whether a new conceptual operation such as weighted combinations of different measurements is necessary or suitable to account for the meanings of certain terms or expressions, and one still has to characterize which terms and expressions are so interpreted. All that a procedural semantics approach does is to increase the inventory of conceptual apparatus that one can utilize in accounting for meanings without requiring each new such conceptual mechanism to be introduced as a new primitive element or undefined concept.

I should point out that I view the procedural approach to the problems of semantics not as an alternative to the more traditional Tarskian model-theoretic account, but rather as a means to supplement that account with what in computer terminology would be called an "upward compatible" extension. That is, I view the Tarskian account as a particular special case of a procedural account - one in which the procedures involved are the definitions of the quantifiers and logical connectives as procedures for assigning truth values to complex propositions as a function of the truth values of their constituents. The Tarskian account, however, stops short of attempting to specify the truth conditions of elementary propositions, and falls short of an adequate account of the various "opaque context" operators

such as "believe" and "want", whose truth values are determined by something more than just the truth values of their constituents. I believe that by viewing the operations of Tarskian model theory as abstract procedures, and making the extension to permit a more diverse range of procedural primitives from which to construct the meanings of utterances, one can obtain an adequate semantic account, not only of propositional utterances, but also of various other speech acts. Moreover, within the same framework, one can extend the range of semantic explanation beyond an account of truth conditions in terms of abstract models (as in a classical model theory) to include an account of how these models relate to the actual world via sensory perception.

12 Some Inadequate Theories of Procedural Semantics

As I stated earlier, there are a number of theories that one could formulate within the procedural semantics framework that are clearly wrong. At this point, I would like to point out two of them, which I will call the "induced effect" theory and the "criteria for belief theory". The first is a theory that whatever procedure the machine carries out in response to a sentence constitutes its meaning. Among other things, this theory would dictate that if a declarative sentence causes some representation of itself to be stored in memory, and if a question causes some searching and matching procedure to be executed to try to find a matching statement in the memory, then these procedures for storage and searching would be the meanings of the sentences in question.

Now there is no question that the effects of many sentences are to invoke such storage and searching operations. What is in question here is whether those operations can sensibly be taken as an explication of the ordinary use of the word "meaning". The answer is that of course they can't, for one quickly realizes that the meaning of a sentence such as "It is raining" really has something to do with whether it is raining, and not whether a representation equivalent to that sentence is stored in one's head (no matter how strongly it might be believed). Thus, the procedures by which a representation of this sentence is constructed and stored will not serve as an explication of its meaning.

The "criteria for belief" theory takes the meaning of a sentence to be the criteria for deciding whether to believe it. While this is somewhat closer to the mark, a little reflection will convince one that the situation is quite the other way around. That is, the criteria for belief may include some appeal to the meaning of a sentence, but this is not the only criterion for deciding what to believe. Other criteria include such things as credibility of the source. One should certainly distinguish such criteria for belief from the meaning of the sentence.

Both of the above "theories" attempt to identify the meaning of a sentence with something that is done when the sentence is understood. In fact, unlike the semantic interpretation procedures, the procedures that characterize the meanings of sentences are usually not executed as part of the understanding of a sentence. Rather, their importance lies in being available for execution and/or for simulated execution in hypothetical situations. They serve to define the standard of what a concept refers to, even when input sentences merely make reference to

such procedures without requiring their actual execution. For example, many assertions can be viewed as statements of a relationship between two procedures, neither of which is executed as part of the process of understanding, but which are nevertheless critical parts of the meanings. Specifically, "Snow is white" can be interpreted as an assertion of a particular relationship (sometimes referred to as "holding") between the meaning function for "white" and the extension of the meaning function for "snow". To understand the meaning of such sentences, it is not necessary to execute these procedures but only to have them.

In addition to avoiding the above two specifically false interpretations of what a procedural account of meaning might be, there are a number of other, somewhat subtle, interpretations of the notion of procedure that need to be made if one is to give an adequate procedural account of the semantics of English. I will argue that in order to adequately model our pretheoretic notion of meaning, one has to depart in several respects from a "straightforward" interpretation of the meanings of expressions as procedures (i.e., as analogous to a piece of computer program that can be executed in any given situation to determine the truth value or referent of the expression in question). Specific problems have to do with the extent to which the procedures can actually be executed in a given situation, and the issue of what aspects of the specification of a procedure in some representation count as essential parts of the meaning. In the next few sections, I will discuss a number of alterations to the "straightforward" notion of procedural semantics that need to be made in order for such a notion to be an adequate theory of meaning. Many of these issues have been touched upon in a previous paper [Woods, 1973].

Before proceeding further, let us first deal with an often raised question relating to the possibility of achieving the goal.

13 A Fallacious Criticism of the Procedural Approach

A long standing criticism of any attempt to formally set out the truth conditions for ordinary terms such as "dog", "chair", and "alive" is that, although various people have attempted such definitions, no one has succeeded in giving any that are satisfying. I think that an argument I will give shortly for the necessity of partial procedures that do not attempt to account for all of the pathological entities which are "neither fish nor fowl" accounts for a major component of such failures. A second component, I believe, is a rather limited view of the possible devices out of which such a definition could be constructed. In particular, one argument for the impossibility of characterizing the meanings of such terms in terms of more primitive properties goes roughly as follows:

If one attempts to characterize the meanings of ordinary terms such as "dog" as having specific definitions in terms of more primitive properties such as having four legs, hair, certain kinds of teeth, eating meat, etc., then in almost all such cases one can imagine (or actually encounter) entities to which the term should apply, but which fail to have one or more of these properties. One can imagine dogs with three legs (either as a result of a physical accident or congenital abnormality), without hair (singed or shaved or somehow bald), with no teeth, etc. It seems that none of the supposed defining properties are

absolutely necessary. Moreover, if there were some absolutely necessary properties (animate? - no, a dead dog is still a dog), their combination would be extremely unlikely to characterize the desired meaning. That is, even if there were some absolutely necessary conditions for being a dog, their combination would not be sufficient to define what a dog is, but only some much larger class of entities.

The fallacy in the above argument is an implicit assumption that the way that the defining properties are combined to make the definition is by simple conjunction. Such implicit assumptions can easily be (mis?)read into discussion such as the following [Lyons, 1977].

...we can define a class on the basis of some property (or set of properties) which they [sic] have in common. Suppose, for example, we summarize the set of properties assumed to be essential for something to qualify as a dog... Then we can say that the class of dogs comprises all those objects in the universe that have this, no doubt very complex,... set of properties. ... the intension of a term is the set of essential properties that determines the applicability of the term.

However, it is not the case that the only way to define a class of objects in terms of properties is to conjoin them. In particular, a procedural combination of elementary properties could involve conditional checking of some properties dependent on the values of others, as in the following hypothetical example (in a hopefully intuitive procedural language), where A, B and C are assumed to be tests of elementary properties:

if A and B then conclude true;
else if not A and not B, conclude false;
else if C, conclude true;
else conclude false.

In this definition, none of the properties A, B, or C are necessary properties for truth, but they are nevertheless the defining properties.

The above example could of course be represented as a disjunction of conjunctive cases without recourse to a notion of procedures. However, procedures in general permit convenient specification of such operations as testing whether the number of properties from a set of possible ones is greater than some threshold, whether the number of confirming properties exceeds the number of disconfirming ones, computing weighted sums of some kind of measures of importance of properties, and applying functions to values of continuous parameters such as height and weight. In such formulations, the connection between some defining property and the ultimate truth value that would be assigned to a proposition could be quite remote and intricate.

Arguing impossibility on the basis of an implicit assumption that limits one's ability to achieve the supposedly impossible goal is a very easy trap to fall into. Even as sophisticated a proceduralist as Winograd appears to be prone to a form of this fallacy (presumably in weak moments). In his account of the previously mentioned "bachelor" example, he says, "In normal use, a word does not convey a clearly definable combination of primitive propositions, but evokes an exemplar ..." [Winograd, 1976]. Here, "clearly definable combination of primitive

propositions," is apparently blocked from including such a clearly defined combination as a pattern match with some matching criterion against an exemplar.

One of the advantages of the procedural approach to semantics (as Winograd is taking pains to point out in the above quote) is that procedures do provide for the clear definition of a concept in terms of more basic properties in ways other than simple conjunction, and (in particular) in ways involving such things as testing whether the number of properties an object has in common with some prototype is above some threshold. Moreover, it permits generalizations of such comparisons to comparisons with several alternative prototypes, selectively counting some properties as more criterial than others, and assigning various notions of approximate truth or degrees of satisfaction of a pattern. One of the major advantages of the procedural semantics paradigm, from my point of view, is that it permits such a range of devices with which to attempt to account for the meanings of words and utterances and their use in language.

14 Partial Functions

One of the first adaptations that must be made to the straightforward procedural semantics account is to realize that one must in general permit meanings to be defined by partial functions that in some cases assign neither true nor false. The necessity for such functions is most strongly motivated by predicates such as "Sentence x is false" from which one can construct sentences that cannot be given any consistent truth value. For example, one of Russell's paradoxes (a version of the

"Liar's paradox" of Epimenides) consisted of writing on one side of a piece of paper "The sentence on the other side of this paper is true" and writing on the other side "The sentence of the other side of this paper is false". Neither of these sentences can be assigned either of the values true or false without thereby inducing a logical inconsistency, although either sentence by itself seems meaningful and can be true or false for other possible values of "the sentence on the other side of this paper".

The problem with this pair of sentences is partly due to the fact that their truths are purportedly mutually defined in terms of each other with no other foundation (i.e., the definition is genuinely circular). Notice that if the pair of sentences both said "The sentence on the other side of this paper is true", either the assignment of both true or both false would be consistent, but there is no principle for choosing one or the other. Worse still, if both sentences were "The sentence on the other side of this paper is false", then the assignment of true to one and false to the other is consistent, although by symmetry one would expect that both sentences should get the same truth value. It seems necessary, therefore, that if we are to give any semantics at all to ordinary English, it must admit the possibility of certain predicates being defined by partial functions that fail to assign truth values at all in some cases (or equivalently have a "third truth value" which is neither true nor false).

The use of partial functions as meaning criteria, once one has been forced to permit it, solves another troublesome problem in the semantics of ordinary terms such as "chair", "dog", and "alive". Unlike formally defined terms like "bachelor", these

ordinary terms do not have a straightforward definition in terms of other words. Instead, they are learned by induction from our experience. As mentioned above, actual attempts to formally characterize the meaning of such terms seem always to fall short of completely delineating the class of objects to which one would want to apply the term. This is claimed as a demonstration of the impossibility of characterizing the meanings of such terms by such means. However, another possible interpretation of the data is that the criteria for the meaning of the term that people have in their heads is in fact partial, although capable of extension to resolve new unanticipated cases. Some psychological experimentation lends credence to this interpretation, since it can be demonstrated that people have difficulty deciding whether to assign the term "chair" to various chair-like objects that violate one or more of the ordinary defining characteristics of "chairhood" (e.g., the absence of a back) [Miller & Johnson-Laird, 1976].

A possible account of these psychological results is that the meaning function for "chair" is in fact a partial procedure that assigns truth in some cases and falsity in others, but has simply never been extended to cover all of the possible sensory stimuli that it could be given as arguments. In this view, what goes on when such a novel instance of chair-like object is presented is that a kind of problem-solving activity is invoked to determine whether to extend the meaning of the term to include this new kind of object or not. These problem-solving processes consider such factors as similarity of the candidate object to various prototypes, the severity (on some scale) of the violations of formerly necessary conditions, the risk of overgeneralization, the utility of the resulting extended meaning, consistency with already held beliefs about classes of

objects, estimates of how other people would use the term, etc. In this view, the process of extending the meaning of a term or sharpening its discrimination is a creative act that is voluntarily taken by a person in certain circumstances. Moreover, if we look at the way that people acquire meanings, it seems clear that some such processes are essential. The meanings of most terms are acquired by extensive exposure to examples of their use and a gradual induction of their meaning. Moreover, those few terms that aren't acquired this way are ultimately defined in terms of ones that are.

If the above view is correct, then the notion of partially defined meaning function is not a strange anomaly of certain abstract theoretical concepts, but rather a ubiquitous characteristic of the meaning of words.

15 Accessibility

Another point of subtlety required for an adequate procedural account of meaning has to do with the question of applicability of the procedures in worlds where the system does not have privileged access to all facts in the world, but must deal with the potential difference between what is true in the world and what the system thinks it knows. Current applications of procedural semantics in systems such as the LUNAR system and Winograd's SHRDLU do not face such problems, since, by definition, the meanings of their expressions refer to the states of their internal computerized models. These systems thus have privileged access to the true state of their worlds in a way that humans (and mobile robots) do not. In the artificial worlds of

these two systems, the primitive "perceptual" routines that measure states of the world not only do not make errors, but also have complete and total access to everything that is true of their world.

There are a number of artifacts of a procedural semantics for such completely accessible worlds that do not extend to the situation in which human beings find themselves in the real world. The most notable of these is the ability to actually execute the procedure that defines the meaning of an expression. In the real world, such procedures frequently are not executable due to lack of access to some of the data on which they operate. For example, even if one has a well-defined procedural specification of the meaning of a color term in terms of the output of a spectral analyzer (say) (or in terms of the sensors in one's eye), one can fail to be able to execute this procedure in practice because there is no light, or the object in question is inside a locked box and cannot be seen, or it is located in China or on the far side of the moon and one cannot get there, etc. Most seriously, the object may have existed in the past (and may now be disintegrated into dust and dispersed to the winds), or it may be a predicted future object, in which case there is no way even in principle (at least within our current technology and beliefs about what is possible) to carry out the procedure.

Nevertheless, even in this most extreme case of inaccessibility of the data on which the procedure would have to be operated, the procedure itself still seems to be a suitable entity to take as the characterization of the meaning of the term. Even in these cases, the representation of the procedure as a structured entity can serve as a source of inferences about

what its outcomes would have been in certain circumstances. For example, the truth conditions for (the most common sense of) the word "bachelor" involves the quantification over moments in past time to determine that the individual in question has never been the groom in a marriage ceremony. This is a perfectly well-defined procedure in spite of the fact that the relevant perceptions in past time, if they were not made at that time, are not available for retrospective testing in the present. One can, and does, of course look for evidence at a later time of the truth of a given fact at an earlier time, but this is not what characterizes the meaning or truth conditions of assertions about the past (since the failure to find such evidence does not imply the falsity of assertion and fraudulent evidence could be planted).

16 Idealized Access

It seems, then, that the English language permits us to talk about quantification over moments of past and future time, although our access to these moments in time is severely constrained. Likewise, it permits us to talk about arbitrary points in space, where our access (while less limited than in the case of time) is also constrained. For example, the language permits us to talk about such things as whether certain structures are present inside human cells, even though the actual perception of such cells and their contents requires technical augmentation via a microscope and prior to the invention of the microscope was not possible at all. Nevertheless prior to the invention of the microscope, the theory of the human body being composed of cells could have been described and understood by

means of an abstract notion of focusing one's attention on smaller and smaller scale in this abstract model of space and time, without specific mention of a practical method for actually gaining such access.

In a similar way, our use of words such as "believe" and "want" to apply to other people seems to apply to an idealized world in which we can focus our attention on the beliefs and perception of others, directly perceiving their internal mental events. Again, this is not possible in practice (with present technology), but appears to be what we mean by many English locutions.

As a consequence of these accessibility limitations, it is clear that if procedures are to be taken as explications of meanings, one cannot expect to just blindly execute them. Rather, in some (most?) cases, an intelligent inference component is required in order to deduce useful information from the procedural specification. This in turn dictates that the procedural specifications must be useful for more than just execution as "black box" procedures with input-output conditions. They must have internal structure that is accessible to inferential procedures.

17 Knowing How versus Knowing Whether

There is another artifact of the procedural semantics used in LUNAR and SHRDLU that will not extend to the general situation. This is the ability to treat the meanings of commands as procedures for carrying them out, and the meanings of nouns as procedures for enumerating the members of the corresponding

class. (For example, the meaning of "rock" to LUNAR is a procedure for enumerating all of the lunar samples in its data base, and the meaning of PRINTOUT is a procedure for printing out answers.) Although such procedures are useful when one has them, they are too strong to demand as a criteria for meaning. In the general case, the meaning for a nominal concept must be something weaker - something that can tell an instance of the concept when it is presented, but may not be able to find or enumerate all (or even any) instances. (In real life, one can recognize a rock when one sees one, but could hardly begin to enumerate all existing ones.) In a similar way, the meaning function for an imperative sentence must be something weaker than a procedure for actually carrying it out, since one can perfectly well understand the meaning of sentences such as "open the box" even when the box in question has a trick latch and one cannot figure out how to open it. Thus, the meaning function for (the propositional content of) an imperative seems to be something like a procedure for recognizing its successful completion rather than actually carrying it out.

There is no questioning the utility of having a procedure for a given verb that knows how to carry it out or cause it to be true, or the utility of having a procedure for finding instances of a given noun. These procedures, however, are practical skills, not criteria for the meanings of the terms with which they are associated.

18 Meaning Functions versus Recognition Functions

Although one would at first expect that the procedure

defining the meaning of a concept would be the procedure that one would use to recognize instances of that concept, because of economic considerations and logical necessities (such as the previously discussed accessibility limitations), this is not usually the case. That is, the procedure that we normally use for determining whether something we perceive is an instance of a concept involves considerably less than checking out the full procedure that defines the meaning of the concept. It is this fact that permits us to be fooled by objects that are not what they appear to be. For example, the meaning of "telephone" requires more than just the appearance of a telephone, since if I examine what appears to be a telephone and discover that it has no mechanism inside it, but is instead filled with plaster, I will not consider it to be a telephone. Nevertheless, if I have not so examined it, I will treat it in every respect as a telephone until I discover the inconsistency. (Even knowing the inconsistency, I may still refer to it as a telephone in communication with someone else, but in such a case, it may be another sense of the word that I will be using.)

One must assume then that concepts in the system's taxonomy will in general have two associated procedural functions - a meaning function and a recognition function. The meaning function defines the "bottom line" truth conditions for the concept, but in general may be difficult or impossible to execute in practice. The recognition function is the procedure that we ordinarily use to determine or estimate the applicability of a concept in practical situations, although this procedure may be fooled. The meaning function constitutes the criterion with respect to which the recognition function is calibrated for reliability. For example, although the meaning function for vertebrates may involve something like dissection, one ordinarily

recognizes an animal as a vertebrate from external visual characteristics (e.g., identification of the animal as a member of a known class). The validation of the use of this recognition function consists of dissection experiments (or something equivalent) that justify the assertion that those recognition characteristics imply the truth of the meaning function (apparently the converse is not required). A methodological test for distinguishing a meaning function from a recognition function is to consider under what circumstances one would admit to having been fooled about whether a concept was satisfied by an entity. In such situations, the recognition function has been satisfied, but the meaning function has not.

Note: I expect that it is psychologically possible for people to use a term for which they have induced a recognition function or part of one, but for which they have no meaning function or only the most nebulous idea of what the meaning function is (e.g., they have only some metabeliefs about the nature of the meaning function). Moreover, it would follow from this that it is possible for an entire culture to make use of a term whose meaning function has not been adequately characterized by anyone, and for which there is no criterion for calibration of the recognition function. In particular this could happen when everyone is under the assumption that there is such a meaning function, although they don't fully understand it (and even when some people think they do fully understand it). It is probably even possible for this to be done when there is something fundamentally inconsistent in the metabeliefs about the meaning function that people hold.

19 Abstract Procedures

Still another subtlety that intrudes on a straightforward procedural account of meanings in English has to do with the level of detail that we wish to have considered as part of the meaning of a term. If one considers, for example, the term "vertebrate", then the procedure involving its definition would involve something like gaining access to the interior of the animal and seeing whether there is a backbone (as by dissection). However, to the extent that this procedure adequately characterizes the meaning of the term, it should not include details such as how the incision is made, what kind of scalpel is used, or the position or angle of the head in looking into the opening. To serve as an adequate model of meaning, such procedures will have to be expressed as very high-level programs that specify subtasks in general terms without commitment to details. There is an increasing tendency in high-level computer programming languages toward exactly this kind of abstraction, so that programmers need specify only the essential characteristics of what is to be done without specifying details that don't make a difference to the outcome.

The kind of abstract procedures required here may be thought of as programs whose subroutines may have alternative procedural realizations (with respect to which the calling function "doesn't care" which is used). A particularly simple example of this kind of don't care condition would be the order in which different clauses of a conjoined condition were tested. Again, the traditional Tarskian account already provides us with a notion of an abstract procedural operation for conjunction in which the order of testing the conjuncts is not considered relevant

(although any given axiomatization will essentially specify an order or set of possible orders in which to do things). In exactly the same way, we can think of the statement of an AND operation in a procedural definition not as a single ordered sequence of tests (the way it would be implemented in most programming languages), but rather an abstract specification of a set of alternative possible orders in which those conditions could be tested.

It is a non-trivial undertaking to construct a suitable procedural expression language that clearly indicates what details are important, and what are irrelevant. In fact, this is one of the current research goals of one segment of programming language theorists. However, it is not too difficult to characterize an abstract semantics for such procedures as essentially corresponding to an equivalence class of more detailed procedures, any one of which will serve as a specification of the truth conditions.

Working out the details of what counts as a significant difference between two procedural representations at the level of abstraction that one would like to use for characterizing meaning functions will probably be a difficult task. Moreover, it will involve a heavily empirical component - attempting to characterize and formalize people's pretheoretic intuitions about sameness of meaning. I suspect that the level of abstraction that turns out to be satisfactory will have something to say about the structure of knowledge representation in human memory (and/or vice versa). Future work in knowledge representation structures will hopefully develop some insights into these issues.

20 Abstract Procedures as Intensional Objects

We have now argued against two extreme interpretations of procedural semantics - a black box approach in which the internal structure of a meaning function was inaccessible (only the input-output relations are available), and a low-level detail approach in which every detail of the operation of the meaning function procedure is considered a "part of the meaning". The former gives rise to a sense of equivalence between meaning functions that is too weak (in the sense of strong and weak equivalence of formal language theory) because it counts as equivalent meaning functions whose input-output relations are the same (in all possible situations) regardless of the means by which those extensions are determined. As a simple example of the consequences of this kind of weak equivalence, the propositions "either the moon is made of green cheese or it isn't" and "Sir Walter Scott is Sir Walter Scott" would mean the same thing since they are both tautologies.

The low level detail interpretation is at the opposite extreme of this spectrum. Its sense of equivalence is so strong that it counts two meaning functions as different if they differ in any detail of their operation regardless of the extent to which they effectively do the same thing. The notion of abstract procedure that is required for the characterization of meaning functions appears to lie somewhere between these extremes - providing a degree of internal structure that is considered significant, while leaving certain low-level details unspecified (or specified with suitable "don't care conditions").

For reasons similar to the above, Carnap [1964] introduced the notion of the "intension" of a predicate to serve as the

thing that characterizes the truth conditions of the predicate but also contains some "intensional" structure beyond that possessed by an abstract set of input-output pairs. (He referred to the latter as the "extension" of the predicate.) In these terms, our "black box" account can be thought of as a kind of extensional account, while our notion of abstract procedure can be taken as an attempt to explicate the notion of intension.

21 Towards a Theory of Meaning Functions

In light of the previous discussions, one can now begin to outline a theory of meaning within the procedural semantics framework that I believe might be viable as an adequate theory. To begin with, it would assume that the meanings of terms are defined as abstract procedures built upon a basic set of perceptual primitives that are essentially those of our own direct perceptions (including internal perceptions of beliefs, desires, emotional states, etc.), but are treated as if these primitives could be applied in arbitrary contexts of time, space, and perceiver.

On top of these primitives are built more abstract predicates, propositions, and functions by use of the compositional operators of some universal machine (in the Turing machine sense) such as recursive function theory, Post production systems, or a modern high level programming language such as LISP. Such a foundation will permit one to construct meaning functions that take into account such factors as numbers of features shared with some prototype, differential diagnosis between two similar concepts, probabilistic calculations,

thresholding decisions, etc., as well as the simple combination of truth values by means of logical operations such as AND and OR and universal and existential quantification.

These composite procedures are not simply black boxes (or abstract sets of input-output conditions), but rather have an intensional structure that permits the intelligent system not only to execute them against the external world in particular situations of time and place (with the system itself as perceiver), but also to simulate them in hypothetical situations, including situations involving other perceivers and/or times and places that are not available to direct perception.

These procedures are abstract in the sense that they are expressed at a level of abstraction that "hides" (or declares non-essential) certain low-level details of operation with respect to which two procedures that otherwise compute the same thing are considered the same.

There are associated with a given term two fundamentally different procedures - a meaning function and a recognition function. The recognition function is the function that is normally used to recognize instances to which the term applies, while the meaning function is the standard against which the recognition function is measured for reliability, and from which (in certain theoretical situations) possible recognition functions can be derived.

Given such a set of notions to work with, it now becomes possible to talk about meanings that are in some sense anchored to actual perceptual operations via an assembly of recursively defined procedural specifications, without the meaning function necessarily being executable in practice in all situations. In

particular, if we want to characterize the meaning of some past tense statement that would otherwise be a directly perceptible fact were it expressed in the present tense, then the assertion has the effect of claiming that if the corresponding procedure had been executed at that time, then its computed value would have been true. This can be expressed abstractly in terms of abstract time and place setting operations followed by an evaluation of the procedure in that abstract setting.

22 Conclusions

In the previous discussion, I have argued that a notion of procedural semantics can serve as an upward compatible generalization of the Tarskian semantics of truth conditions, and that specifically it provides natural extensions of similar techniques to kinds of meanings that are not directly expressible as having truth conditions. Moreover, it permits the extensions of semantic theories to account for the way in which the truth conditions of elementary "atomic" propositions connect to our perceptual experiences. However, I have raised a number of issues with respect to which the most natural and straightforward notions of procedural semantics are inadequate, and have outlined the direction in which I think an adequate solution lies.

I believe there are good methodological reasons to adopt a procedural semantics approach to the characterization of meaning. Specifically, there now exists a sizeable body of intuitive understanding of the nature of procedures, sharpened by the rigors of making a machine actually perform as intended in response to procedural specifications - especially to make a

machine perform (albeit in limited ways) tasks normally thought of as higher level mental processes (parsing sentences, answering questions, proving theorems, etc.). I believe these intuitions and insights are invaluable in stretching one's view of what kinds of internal mental processes are possible. Moreover, theoretical results in automata theory and computability give a depth of theoretical understanding to some of the issues that arise that are not available for less concrete notions of propositions, predicates, intensions, etc.

It's not that the notion of a procedure is admirably well-suited to this kind of analysis. Rather it seems that there is no other mathematical entity as well understood as that of a procedure that one can use to construct a more adequate explanation of the phenomena. Calling such things "propositions" does not help, since we have no independent explication of what a proposition is. Calling them functions (in the abstract mathematical sense of sets of ordered pairs) is merely a black box account that refuses to deal with the contents of the black box (i.e., how the function in fact assigns values to arguments).

However, one should be aware that the notion of procedure, while somewhat more concrete than the abstract notion of proposition, is nevertheless a fairly subtle concept in its own right, and its invocation as a mechanism for modeling semantics is not as straightforward as it might first appear. The final point that I would like to make is that adopting a procedural semantics approach does not automatically provide a solution to all of the classical problems and paradoxes of semantics. Saying that the meanings of English expressions are abstract procedures does not eliminate the need for concern over how such procedures are represented, and how expressions of similar meaning share

common meaning elements. It merely answers the question of what kinds of things these representations can be interpreted to be and what functions they serve.

Much remains to be done to work out an adequate model of meaning. In a sequel to this paper I intend to discuss a number of difficult problems of semantics from a procedural perspective. Some of these problems (opaque contexts, presuppositions, the uniqueness of identity, and theoretical concepts) are traditional problems in the philosophy of language. Some others (abstract procedures, infinite quantification) are problems unique to the procedural approach.

In advancing our understanding of human cognition (just as in Dennett's account of intelligence), there seem to be two important components - hypothesizing models (or aspects of models) and developing criteria by which such models are to be judged. In this paper, I hope that I have accomplished a little of both.

Acknowledgments

The first draft of a preliminary version of this paper was written while on sabbatical at the University of Sussex, under the auspices of a Fulbright-Hays Fellowship. Subsequent development of these ideas was supported in part by ONR under contract N00014-77-C-0371 and by ARPA under contract N00014-77-C-0378. Special thanks are due to Christopher Longuet-Higgins and Stuart Sutherland, who provided the initial environment for this work, Mary Walton, who heroically (and with exceptional skill) typed the first draft under great pressure, and to many people who read and commented on the draft, including especially Eugene Charniak, Daniel Dennett, Gerald Gazdar, Steven Isard, David Israel, Philip Johnson-Laird, and Aaron Sloman (none of whom can be held responsible, of course, for the opinions expressed herein). Special thanks are also due to John Lyons for valuable discussions.

REFERENCES

- Bell, A. and Quillian, M.R. 1971. "Capturing Concepts in a Semantic Net," in E.L. Jacks (ed.), *Associative Information Techniques*. American Elsevier Publishing Co., Inc.
- Berkeley, E.C. and Bobrow, D.G. (eds.). 1964. *The Programming Language LISP: Its Operation and Applications*. Cambridge, MA: MIT Press.
- Brachman, R.J. 1978. "A Structural Paradigm for Representing Knowledge," Ph.D. dissertation, Division of Engineering and Applied Physics, Harvard University. (Also, BBN Report No. 3605, Bolt Beranek and Newman Inc., Cambridge, MA, December. Now available from NTIS as AD No. AO56524.)
- Brachman, R.J. 1979. "On the Epistemological Status of Semantic Networks." In *Associative Networks - The Representation and Use of Knowledge in Computers*, Nicholas V. Findler (ed.). New York: Academic Press.
- Carnap, R. 1964. *Meaning and Necessity*. Chicago: University of Chicago Press.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge: M.I.T. Press.
- Church, A. 1964. "The Need for Abstract Entities in Semantic Analysis," in J.J. Katz and J.A. Fodor (eds.), *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall.
- Dennett, D.C. 1974. "Why the Law of Effect Will Not Go Away." Read to the first meeting of the Society for the Philosophy of Psychology, October 26, 1974 at M.I.T. Reprinted in *Brainstorms: Philosophical Essays on Mind and Psychology*, Bradford Books, 1978.
- Frege, G. 1892. "Über Sinn und Bedeutung". *Zeitschr. f. Philosophie und philosoph. Kritik* 100. 25-50. English translation: "On Sense and Reference," in Geach and Black (1960: 56-78). Reprinted in Zabeeh et al., 1974: 118-40; Feigl and Sellars, 1949: 82-102; etc.
- Katz, J.J. and Fodor, J.A. 1964. "The Structure of a Semantic Theory," in Katz and Fodor (eds.), *The Structure of Language: Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall.

- Lyons, J. 1977. Semantics:1. London: Cambridge University Press.
- Miller, G.A. and Johnson-Laird, P.N. 1976. Perception and Language. Cambridge, MA: Harvard University Press and London: Cambridge, University Press.
- Montague, R. 1970. "English as a Formal Language," in Bruno Visentini et al. (eds.), *Linguaggi nella societa e nella tecnica* (Proc. of a symposium, "Language in Society and the Technical World," held in Milan in Oct. 1968), Milan: Edizioni di Comunita, pp. 189-224.
- Montague, R. 1973. "The Proper Treatment of Quantification in Ordinary English," in J. Hintikka, J.M.E. Moravcsik, and P. Suppes (eds.), *Approaches to Natural Language*. Dordrecht, Holland: Reidel.
- Quillian, M.R. 1966. "Semantic Memory". Report No. AFCRL-66-189, Bolt Beranek and Newman, Inc., Cambridge, MA 02138, October (more extensive than the Quillian 1968 paper by the same name).
- Quillian, M.R. 1968. "Semantic Memory," in M. Minsky (ed.), *Semantic Information Processing*. Cambridge, MA: MIT Press.
- Quillian, M.R. 1969. "The Teachable Language Comprehender: A Simulation Program and Theory of Language," in *Communications of the ACM*. 12, pp. 459-476.
- Tarski, A. 1944. "The Semantic Conception of Truth," in *Philosophy and Phenomenological Research* 4, pp. 341-375. (Reprinted in Tarski, 1956; Olszewsky, 1969; Zabeeh et al., 1974).
- Weissman, C. 1967. LISP 1.5 Primer Belmont, CA: Dickenson Publishing Co.
- Winograd, T. 1972. *Understanding Natural Language*. New York: Academic Press.
- Winograd, T. 1976. "Towards a Procedural Understanding of Semantics," Memo AIM-292, Stanford Artificial Intelligence Laboratory, November.
- Woods, W.A. 1967. "Semantics for a Question-Answering System," Ph.D. thesis, Division of Engineering and Applied Physics,

- Harvard University. Also Report NSF-19, Harvard Computation Laboratory, September. (Available from NTIS as PB-176-548, and reprinted with a new preface in 1979 by Garland Publishing, Inc.)
- Woods, W.A. 1970. "Transition Network Grammars for Natural Language Analysis," CACM, Vol. 13, No. 10, October.
- Woods, W.A. 1973. "Meaning and Machines," in Computational and Mathematical Linguistics, A. Zampolli (ed.), Proceedings of the International Conference on Computational Linguistics, Pisa, Italy, August, 1973. Leo S. Olschki, Florence, Italy, 1977.
- Woods, W.A. 1978a. "Semantics and Quantification in Natural Language Question-Answering," in Advances in Computers, Vol. 17. New York: Academic Press. (Also Report No. 3687, Bolt Beranek and Newman Inc.; now available from NTIS as AD No. AO47600.).
- Woods, W.A. 1978b. "Taxonomic Lattice Structures for Situation Recognition," in TINLAP-2, Conference on Theoretical Issues in Natural Language Processing-2, University of Illinois at Urbana-Champaign, July 25-27. (Also in AJCL, Mf. 78, 1978:3).
- Woods, W.A. and Brachman, R.J. 1978. Research in Natural Language Understanding, Quarterly Technical Progress Report No. 1, 1 September 1977 to 30 November 1977, Report No. 3742, Bolt Beranek and Newman Inc., Cambridge, MA, January. (Now available from NTIS as AD No. AO53958.)
- Zadeh, L.A. 1974. "Fuzzy Logic and Approximate Reasoning." Electronics Research Laboratory Memorandum No. ERL-M479, College of Engineering, University of California, Berkeley.

Official Distribution List
Contract N00014-77-C-0371

	<u>Copies</u>
Defense Documentation Center Cameron Station Alexandria, VA 22314	12
Office of Naval Research Information Systems Program Code 437 Arlington, VA 22217	2
Office of Naval Research Code 200 Arlington, VA 22217	1
Office of Naval Research Code 455 Arlington, VA 22217	1
Office of Naval Research Code 458 Arlington, VA 22217	1
Office of Naval Research Branch Office, Boston 495 Summer Street Boston, MA 02210	1
Office of Naval Research Branch Office, Chicago 536 South Clark Street Chicago, IL 60605	1
Office of Naval Research Branch Office, Pasadena 1030 East Green Street Pasadena, CA 91106	1
Office of Naval Research New York Area Office 715 Broadway - 5th Floor New York, NY 10003	1
Dr. Martin Epstein National Library of Medicine Bldg. 38A, 8th Floor Lab 8600 Rockville Pike Bethesda, MD 20209	1

Director, National Security Agency 1
Attn: R53, Mr. Glick
Fort G.G. Meade, MD 20755

Naval Research Laboratory 6
Technical Information Division
Code 2627
Washington, D.C. 20380

Naval Ocean Systems Center 1
Advanced Software Technology Division
Code 5200
San Diego, CA 92152

Dr. A.L. Slafkosky 1
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, D.C. 20380

Mr. E.H. Gleissner 1
Naval Ship Research & Development Center.
Computation & Mathematics Dept.
Bethesda, MD 20084

Captain Grace M. Hopper 1
NAICOM/MIS Planning Branch (OP-916D)
Office of Chief of Naval Operations
Washington, D.C. 20350

Mr. Kin B. Thompson 1
NAVDAC 33
Washington Navy Yard
Washington, D.C. 20374

Advanced Research Projects Agency 1
Information Processing Techniques
1400 Wilson Boulevard
Arlington, VA 22209

Captain Richard L. Martin, USN 1
3311 W. Avenue
Newport News, VA 23607

Director 1
National Security Agency
Attn: R54, Mr. Page
Fort G. G. Meade, MD 20755

Major James R. Kreer 1
Chief, Information Sciences
Department of the Air Force
Air Force Office Of Scientific Research
European Office of Aerospace Research
and Development
Box 14
FPO New York 09510

DATE
FILMED
5 8